Evaluating Incidence and Impact Estimates of the Coronavirus Outbreak from Official and Non-Official Chinese Data Sources

Preprint · February 2020

CITATIONS
0

2 authors, including:

Lucia Dunn
The Ohio State University
37 PUBLICATIONS **405** CITATIONS

SEE PROFILE

Evaluating Incidence and Impact Estimates of the Coronavirus Outbreak from Official and Non-Official Chinese Data Sources

February 18, 2020

Mai He, M.D., Ph.D. Associate Professor of Pathology and Immunology, Washington University School of Medicine in St. Louis
Lucia Dunn, Ph.D., Professor Emerita of Economics, Ohio State University, Columbus, OH[1]

INTRODUCTION

The spread of the coronavirus (COVID-19) has fast evolved into a global public health concern. Many key issues remain unknown, such as the origins of this virus, its clinical spectrum, the infection time of the first case, and routes for human transmission. The epidemiological data provide important information for local and global decision making. The Chinese government has been updating such data as the number of confirmed diagnoses and deaths, and the World Health Organization has used some of those data as input in its decision making. The rapid spread of the virus, however, together with reports from the official Chinese media on large numbers of possible coronavirus victims who were not treated within the medical establishment and hence may have fallen outside of the government statistics, has led many to believe that there may be serious gaps in our understanding of the outbreak based on what can be determined from official government data figures.[1] It is therefore important to follow up on all reasonable sources of data, including those outside of normal government channels, to try to accurately assess the number of cases and impact of the outbreak.

The purpose of the current paper is to present credible data sources, including non-official sources, for the medical community and broader public. The truth may lie somewhere in between the extreme values presented here. However, an understanding the range of values coming from credible sources is a critical first step to evaluating any statistical errors in estimates of the epidemic's size and rate of spread.

MATERIALS & METHODS

Data Sources

(a) The official Chinese government figures including Chinese official media reports. We also follow two types of non-official Chinese data: (b) Information as of January 25, 2020, on the operation of crematory facilities in Wuhan [2, 3]. (c) Information on the *Tencent News* website which appeared briefly on two different occasions (January 27, 2020 and February 1, 2020) and

[1]Dr. He and Professor Dunn are part of a team of medical and statistical professionals who have worked on this project.

was reported by *Liberty News* in Taiwan [4]. The data appearing on both of these dates were taken down within hours.

Assumptions

I. Different estimates will be presented, using exponential growth rates corresponding to three different estimates in recent research literature for the epidemic doubling time: (i) 5.8 days [5]; (ii) 6.4 days [6]; (iii) 7.4 days [7]. These numbers apply to the stage when there were no countermeasures in effect.

II. Outcomes are either survival with probability $(1 - d)$ where d is the case fatality rate; or death after 10.5 days with probability d. This number represents an average and has been suggested by medical personnel familiar with the situation. We are unable to determine if there is any consensus on this in the medical literature.

III. To extrapolate from the Wuhan cremation data to the total number of infected cases, alternative assumptions of 2.5%, 5% and 10% were for the case fatality rate.

Methods

Since there is limited information on the details of the transmissibility of the infection, estimates in the tables were calculated from a simple exponential growth model for $n(t)$, the number of new cases per day at time t, where t is the number of days since the occurrence of the initial case. The cumulative number of cases then has the form $N(t) = c\left(e^{\lambda t} - 1\right) + 1$, so that $N(0) = 1$, where c is a constant and the parameter λ is set to match the reported doubling time. The quantities of interest are $N(t)$, $n(t)$, the cumulative number of deaths $D(t)$, the rate of deaths per day, and the case fatality rate d. The model provides relationships between these quantities, such as $n(t) = \left(1 - e^{\lambda}\right)N(t)$ and $D(t) = d\, e^{-10.5\,\lambda} N(t)$.

Under these assumptions, (i) unreported quantities of interest are estimated from available inputs, (ii) quantities are projected from the reporting date to a common reference date (February 12, 2020), and (iii) an approximate starting date (for the initial case) is inferred by extrapolating back from the value of $N(t)$ on the reporting date.

Estimating the starting date requires an estimate of the constant term c. Matching the cumulative totals $N(t)$ to the model where each existing case infects r_0 (the basic reproductive number) new cases on average after s days (the serial interval) gives $c = r_0/(1 - r_0)$. The sources quoted above for the doubling time give estimates 2.3, 2.7 and 2.2 for r_0, and these will be paired with the corresponding doubling time estimates.

The projections of $N(t)$ for Wuhan to the reference date February 12, 2020 (assuming that countermeasures had not been effective by that date) are in some cases large enough (over 10% of the population) that the exponential growth model is unreliable, since it does not take into account the declining number of susceptible individuals. In those cases the model was modified (as in the SIR simulation model) so that the relationship $n(t) = e^{\lambda} n(t-1)$ becomes $n(t) = e^{\lambda}\left(1 - \frac{N(t-1)}{P}\right)^{\frac{1}{s}} n(t-1)$ where P is the population size. For the other data sources, the

exponential growth model was retained since the relevant population is not well defined in those cases.

The 95% confidence intervals presented in the Appendix A are based on the reported CrI (5.8, 7.1) for the doubling rate as presented in [6]. The lower and upper bounds of the corresponding confidence interval for λ were used to compute the confidence intervals in the table. Confidence intervals for the estimated starting dates are not provided because they depend on estimates of both λ and r_0 and also on the randomness in the transmission of the infection in its very early stages.

RESULTS

Crematories in Wuhan, which under normal circumstances would operate about 4 hours per day, were observed to be operating at or close to around-the-clock on January 25, 2020 [2, 3]. This would put the current operating rate at about six times normal. Normal deaths per day can be estimated as 136 based on an annual case fatality rate of 0.00551 in a population of approximately 9 million (Wuhan government data) [8]. Given the capacity of the crematoriums, these additional 20 hours of daily operation imply deaths of (20/4) × 136 = 680 individuals per day above normal.

Tencent briefly reported that there were 213,651 and 233,831 cumulative cases respectively on January 27 and February 1, 2020 [4]. The number of infections and deaths reported on the Tencent website "leaks" are of similar magnitude to those computed from the Wuhan crematory data assuming a 10% case fatality rate.

Calculations based on the middle estimate of a 6.4 day doubling rate (as reported in *Lancet*) are presented in Table 1, and the other two results based on the assumed doubling rates from *NEJM* and *Annals of Internal Medicine* are presented in Appendix B. Results are presented for all scenarios under consideration, including those where the projected number of infections may seem excessively large. The results in Table 1 with confidence intervals included are presented in Appendix A.

The projections to the reference date (February 12, 2020) are valid under the assumption that countermeasures to slow the spread of the epidemic were ineffective up to that date.

Table 1: Data Sources and Estimates for Coronavirus Outbreak

6.4 Day Doubling Rate: (Per *Lancet*, 1/31/2020)

Data Source	Report Date	No. of Cases – Cumulative		New Cases Per Day	Deaths Cumulative	Deaths Per Day	Case Fatality Rate	Implied Start Date
		As of Reported Date	Projected to 2/12/2020					
Chinese Government 1	**2/1/2020**	**33,929**	*112,000*	*3,480*	**304**	*31*	*2.8%*	*11/1/2019*
Chinese Government 2	**2/12/2020**	**73,317**	**73,317**	*7,530*	**1,368**	*140*	*5.8%*	*11/5/2019*
Tencent 1	**1/27/2020**	**213,651**	*1,140,000*	*21,700*	**12,781**	*1,310*	*18.5%*	*10/10/2019*
Tencent 2	**2/1/2020**	**233,831**	*770,000*	*24,000*	**24,589**	*2,520*	*32.8%*	*10/14/2019*
Wuhan Crematory Case 1: 10% case fatality rate	**1/25/2020**	*206,000*	*1,350,000*	*20,900*	*6,650*	**680**	Assume **10%**	*10/8/2019*
Wuhan Crematory Case 2: 5% case fatality rate	**1/25/2020**	*409,000*	*2,510,000*	*41,100*	*6,670*	**680**	Assume **5**%	*10/2/2019*
Wuhan Crematory Case 3: 2.5% case fatality rate	**1/25/2020**	*811,000*	*4,360,000*	*79,800*	*6,720*	**680**	Assume **2.5**%	*9/25/2019*

- Input data is in bold typeface; calculated quantities and dates are in italics. Estimates have been rounded to 3 significant figures.
- New Cases Per Day, Cumulative Deaths, and Deaths Per Day are as of the report date.
- Cumulative number of cases (when input) includes both confirmed cases and suspected cases.
- The Tencent data and Government data refer to all of China, while the crematory data refer to the Wuhan area only.
- Case fatality rate is estimated number of deaths per day on reporting date divided by estimated number of new cases per day 10.5 days earlier.

DISCUSSION

Due to the nature of controls on information in China, none of the data sources presented here could be independently verified firsthand. This applies also to the official Chinese government data, as what would be considered as independent outside verification is not permitted. Based on the number of reported cremations and reports of possibly more accurate numbers on cases that appeared briefly on Tencent and were suppressed, we have conducted a modeling study to estimate the incidence and other parameters of the COVID-19 outbreak, including the potential starting time.

Note first that all sets of non-official data used here imply that the coronavirus outbreak began in October or late September 2019. As reported in the *Lancet* [9], the symptom onset of the first confirmed case was December 1, 2019. Extrapolating back from the official Chinese numbers released February 4, 2020 and February 12, 2020, with the model used here, gives start dates in early November, as noted in the table. This is consistent with other studies [10, 11].

The numbers for cumulative cases, new cases per day, and cumulative deaths coming from the non-official data are all substantially higher by a factor of 5 to 10 (depending on the date of the data) than would be implied by the official Chinese government data. Note that both the crematory data assuming a 10% case fatality rate and the Tencent data give comparable results. Note also that the number of infected cases based on the crematory data will be inversely related to the case fatality rate assumed in their analysis.

The case fatality rates that come from these calculations using the Tencent data are much higher than any that have been recognized by Chinese authorities or the WHO. (For reference, the case fatality rate from the SARS epidemic outside China was approximately 10%.) Note, however, that the case fatality rates inferred from the *Tencent News* data apply only to confirmed plus suspected cases. There are undoubtedly many other cases which may have been misdiagnosed or which never came to the attention of medical personnel, and their inclusion might substantially lower the calculated case fatality rates presented here.

Readers are reminded of the assumptions which underlie these estimates, and therefore they should be taken as approximate. As noted earlier, the true numbers may lie somewhere in between the extreme values reported here. However, even if there were non-negligible reporting errors in these new data, the magnitude of the discrepancy between the results from their analysis and the official figures suggests the need for a serious re-evaluation of all that is known about the current outbreak. In order for the world to respond to this global humanitarian crisis, no credible data source should be ignored.

REFERENCES

[1] https://chinadigitaltimes.net/chinese/2020/02/ 财经-统计数字之外的人：他们死于普通肺炎 [*Were cases excluded from official data thought to be regular pneumonia?*] Last accessed 2/12/2020.

[2] Jiawei Li. The daily death toll in Wuhan exceeds 200. https://www.epochtimes.com/gb/20/1/28/n11828240.htm Last accessed 2/12/2020.

[3] Cathy He. Amid virus outbreak, funeral home officials in Wuhan reveal sharp increase in cremations. https://www.theepochtimes.com/exclusive-funeral-homes-in-coronavirus-ground-zero-cremating-dozens-of-bodies-a-day_3228938.html Last accessed 2/12/2020.

[4] *Liberty News,* Taiwan. https://news.ltn.com.tw/news/world/breakingnews/3056465 Last accessed 2/12/2020.

[5] Ashleigh R. Tuite and David N. Fisman. Reporting, epidemic growth, and reproduction numbers for the 2019 novel coronavirus (2019-nCoV) epidemic. *Ann Intern Med.* February 5, 2020. DOI: 10.7326/M20-0358.

[6] Joseph T. Wu, Kathy Leung, and Gabriel M Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* 2020 Jan 31. PII: S0140-6736(20)30260-9. DOI: 10.1016/S0140-6736(20)30260-9.

[7] Qun Li, Xuhua Guan, Peng Wu, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med.* 2020 Jan 29. DOI: 10.1056/NEJMoa2001316.

[8] Wuhan City Administration. http://www.wh.gov.cn/2019_web/whyw/201910/t20191021_280212.html Last accessed 2/12/2020.

[9] Chaolin Huang, Yeming Wang, Xingwan Li, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020 Jan 15.

[10] Jon Cohen. Wuhan seafood market may not be source of novel virus spreading globally. https://www.sciencemag.org/news/2020/01/wuhan-seafood-market-may-not-be-source-novel-virus-spreading-globally Last accessed 2/18/2020.

[11] Kristian_Andersen. Clock and TMRCA based on 27 genomes. http://virological.org/t/clock-and-tmrca-based-on-27-genomes/347 Last accessed 2/18/2020.

Appendix A: Estimates and Confidence Intervals

Table A: Data Sources and Estimates for Coronavirus Outbreak

6.4 Day Doubling Rate (Per *Lancet*, 1/31/2020)

Data Source	Report Date	No. of Cases – Cumulative		New Cases Per Day	Deaths Cumulative	Deaths Per Day	Case Fatality Rate	Implied Start Date
		As of Report Date	Projected to 2/12/2020					
Chinese Government 1	**2/1/2020**	**33,929**	*112,000 (99,000, 126,000)*	*3,480 (3,150, 3,820)*	**304**	*31 (28, 34)*	*2.8% (2.5%, 3.1%)*	*11/1/2019*
Chinese Government 2	**2/12/20**	**73,317**	**73,317**	*7,530 (6,800, 8,250)*	**1,368**	*140 (127, 154)*	*5.8% (5.2, 6.5%)*	*11/5/19*
Tencent 1	**1/27/2020**	**213,651**	*1,140,000 (969,000, 1,350,000)*	*21,700 (19,600, 23,800)*	**12,781**	*1,310 (1,180, 1,430)*	*18.5% (16.5%, 20.8%)*	*10/10/2019*
Tencent 2	**2/1/2020**	**233,831**	*770,000 (682,000, 869,000)*	*24,000 (21,700, 26,300)*	**24,589**	*2,520 (2,280, 2,770)*	*32.8% (29.2%, 36.8%)*	*10/14/2019*
Wuhan Crematory Case 1: 10% case fatality rate	**1/25/2020**	*206,000 (203,000, 210,000)*	*1,350,000 (1,110,000, 1,650,000)*	*20,900 (18,600, 23,400)*	*6,650 (6,060, 7,360*	**680**	Assume **10%**	*10/8/2019*
Wuhan Crematory Case 2: 5% case fatality rate	**1/25/2020**	*409,000 (405,000, 419,000)*	*2,510,000 (2,090,000, 3,010,000)*	*41,100 (36,700, 46,100)*	*6,670 (6,080, 7,390)*	**680**	Assume **5**%	*10/2/2019*
Wuhan Crematory Case 3: 2.5% case fatality rate	**1/25/2020**	*811,000 (803,000, 828,000)*	*4,360,000 (3,740,000, 5,050,000)*	*79,800 (71,300, 89,200)*	*6,720 (6,120, 7,450)*	**680**	Assume **2.5**%	*9/25/2019*

- 95% confidence intervals in parentheses are based on the reported CrI (5.8, 7.1) for the estimated doubling time.
- Input data is in bold typeface; calculated quantities and dates are in italics. Estimates have been rounded to 3 significant figures.
- New Cases Per Day, Cumulative Deaths, and Deaths Per Day are as of the report date.
- Cumulative number of cases (when input) includes both confirmed cases and suspected cases.
- The Tencent data and Government data refer to all of China, while the crematory data refer to the Wuhan area only.
- Case fatality rate is estimated number of deaths per day on reporting date divided by estimated number of new cases/day 10.5 days earlier.

Appendix B: Estimates Based on Alternate Doubling Rate Assumptions

Table B1: Data Sources and Estimates for Coronavirus Outbreak

5.8 Day Doubling Rate (Per *Annals of Internal Medicine*, 2/5/2020)

Data Source	Report Date	No. of Cases – Cumulative		New Cases Per Day	Deaths Cumulative	Deaths Per Day	Case Fatality Rate	Implied Start Date
		As of Reported Date	Projected to 2/12/2020					
Chinese Government 1	**2/1/2020**	**33,929**	*126,000*	*3,820*	**304**	*34*	*3.1%*	*11/10/2019*
Chinese Government 2	**2/12/2020**	**73,317**	**73,317**	*8,260*	**1,368**	*154*	*6.5%*	*11/15/2019*
Tencent 1	**1/27/2020**	**213,651**	*1,350,000*	*23,800*	**12,781**	*1,440*	*20.8%*	*10/20/2019*
Tencent 2	**2/1/2020**	**233,831**	*871,000*	*26,300*	**24,589**	*2,770*	*36.9%*	*10/25/2019*
Wuhan Crematory Case 1: 10% case fatality rate	**1/25/2020**	*210,000*	*1,630,000*	*23,400*	*6,060*	**680**	Assume **10**%	*10/19/2019*
Wuhan Crematory Case 2: 5% case fatality rate	**1/25/2020**	*418,000*	*2,930,000*	*45,900*	*6,080*	**680**	Assume **5**%	*10/13/2019*
Wuhan Crematory Case 3: 2.5% case fatality rate	**1/25/2020**	*825,000*	*4,800,000*	*88,200*	*6,130*	**680**	Assume **2.5**%	*10/8/2019*

- Input data is in bold typeface; calculated quantities and dates are in italics. Estimates have been rounded to 3 significant figures.
- New Cases Per Day, Cumulative Deaths, and Deaths Per Day are as of the report date.
- Cumulative number of cases (when input) includes both confirmed cases and suspected cases.
- The Tencent data and Government data refer to all of China, while the crematory data refer to the Wuhan area only.
- Case fatality rate is estimated number of deaths per day on reporting date divided by estimated number of new cases/day 10.5 days earlier.

Table B2: Data Sources and Estimates for Coronavirus Outbreak

7.4 Day Doubling Rate: (Per *New England Journal of Medicine*, 1/29/2020)

Data Source	Report Date	No. of Cases – Cumulative		New Cases Per Day	Deaths Cumulative	Deaths Per Day	Case Fatality Rate	Implied Start Date
		As of Reported Date	Projected to 2/12/2020					
Chinese Government 1	**2/1/2020**	**33,929**	*95,100*	*3,030*	**304**	*27*	*2.4%*	*10/19/2019*
Chinese Government 2	**2/12/2020**	**73,317**	**73,317**	*6,560*	**1,368**	*122*	*5.0%*	*10/22/2019*
Tencent 1	**1/27/2020**	**213,651**	*956,000*	*19,100*	**12,781**	*1,140*	*16.0%*	*9/242019*
Tencent 2	**2/1/2020**	**233,831**	*655,000*	*21,000*	**24,589**	*2,200*	*28.1%*	*9/28/2019*
Wuhan Crematory Case 1: 10% case fatality rate	**1/25/2020**	*203,000*	*1,030,000*	*17,900*	*7,640*	**680**	Assume **10**%	*9/23/2019*
Wuhan Crematory Case 2: 5% case fatality rate	**1/25/2020**	*404,000*	*1,930,000*	*35,100*	*7,680*	**680**	Assume **5**%	*9/15/2019*
Wuhan Crematory Case 3: 2.5% case fatality rate	**1/25/2020**	*800,000*	*3,400,000*	*67,800*	*7,760*	**680**	Assume **2.5**%	*9/8/2019*

- Input data is in bold typeface; calculated quantities and dates are in italics. Estimates have been rounded to 3 significant figures.
- New Cases Per Day, Cumulative Deaths, and Deaths Per Day are as of the report date.
- Cumulative number of cases (when input) includes both confirmed cases and suspected cases.
- The Tencent data and Government data refer to all of China, while the crematory data refer to the Wuhan area only.
- Case fatality rate is estimated number of deaths per day on reporting date divided by estimated number of new cases/day 10.5 days earlier.